Information Exchange Limits in Cooperative MIMO Networks

Ali Tajer[*]   Xiaodong Wang [†]

Abstract

Concurrent presence of *inter*-cell and *intra*-cell interferences constitutes a major impediment to reliable downlink transmission in multi-cell multiuser networks. Harnessing such interferences largely hinges on two levels of information exchange in the network: one from the users to the base-stations (*feedback*) and the other one among the base-stations (*cooperation*). We demonstrate that exchanging a *finite* number of bits across the network, in the form of feedback and cooperation, is adequate for achieving the optimal capacity scaling. We also show that the *average* level of information exchange is independent of the number of users in the network. This level of information exchange is considerably less than that required by the existing coordination strategies which necessitate exchanging *infinite* bits across the network for achieving the optimal sum-rate capacity scaling. The results provided rely on a constructive proof.

1   Introduction

Multiple-antenna broadcast transmission is known to offer substantial improvement in spectral efficiency via spatial multiplexing. Such gains depend heavily on the availability of some channel state information (CSI) at the transmitter. Acquiring such CSI allows for managing the intra-cell interference which is due to spatial multiplexing. There exists a rich literature on the design of user scheduling that spans a wide range of transceiver designs with different levels of complexity and CSI feedback requirements [1, 2, 3].

In *multi*-cell multiuser networks, besides the intra-cell interference caused by the residual interference remaining after spatial multiplexing within each cell, another impediment arises from inter-cell interference due to the ever-shrinking cell sizes, which makes the networks operate in interference-limited, rather than noise-limited, regimes. Alleviating the effects of inter-cell interference requires the base-stations to cooperatively adjust their transmissions [4, 5, 6, 7]. Coordination

---

[*]Electrical Engineering Department, Princeton University, Princeton, NJ 08540 (email: tajer@princeton.edu).

[†]Electrical Engineering Department, Columbia University, New York, NY 10027 (email: wangx@ee.columbia.edu).

among the base-stations for downlink transmission imposes some information exchange (cooperation) among the base-stations. Once *full* CSI feedback and *full* cooperation is viable, i.e., *all* the users report their CSI to *all* base-stations, and *all* base-stations share their information streams, the multi-cell network can be essentially modeled as a single-cell multiple-input multiple-output (MIMO) broadcast channel with a transmit antenna array of a higher dimension. Using the literature on MIMO broadcast channel immediately establishes the optimal sum-rate capacity scaling of multi-cell multiuser networks which is achievable via dirty-paper coding [3].

Such optimal capacity scaling, however, necessitates *full* feedback and cooperation across the network. This requirement, specially in large networks, can potentially overburden the feedback channels as well as the backbone channels that connect the base-stations. Therefore, it becomes imperative to study whether such full information exchange is all necessary or contains some redundancy that can be avoided. The question we are specifically trying to address is: what is the minimum amount of information exchange, in the form of either CSI feedback or base-station cooperation, that ensures retaining the optimal sum-rate capacity scaling?

We consider a multi-cell network model, in which every few neighboring cells are collectively regarded as a super-cell that serve the users within that super-cell collaboratively. The analysis demonstrates that for the purpose of achieving the optimal scaling, cooperation among the super-cells, while desirable, is not necessary. In other words, while collective transmission by the base-stations across the super-cells has certain advantages, as long as achieving the optimal capacity scaling is concerned, the super-cells can decide about their scheduling policies independently. Furthermore, each user has to provide feedback to only the base-stations of its designated super-cell. These two relaxations essentially imply that base-stations of each super-cell collect only *local* information and any information exchange across super-cells are avoided. We also quantify the aggregate amount of information exchange necessary for achieving the optimal sum-rate capacity scaling. Specifically, we show that when the network contains M super-cells, each consisting of Q base-stations and when each base-station is equipped with N_t transmit antennas, the minimum aggregate amount of information exchange in the network to guarantee optimal capacity scaling is upper bounded by $MQN_t\lceil \log QN_t \rceil$, which is independent of the number of users in the network.

These results in this paper are provided through a constructive proof. The flow of the proof is as follows. After discussing the network model in Section 2 we describe a scheduling procedure and feedback mechanism in Section 3. Next, in Section 4 we show that proposed scheduler yields the optimal sum-rate capacity scaling. Then in Section 5 we assess the aggregate amount of information

exchange that the proposed scheduler requires. Section 6 addresses some practical issues, e.g., the impact of the size of the super-cells and fairness, and finally Section 7 concludes the paper. To enhance the flow of the content, most of the proofs are relegated in the appendices.

We remark that as opposed to multi-cell MIMO networks, there exists a rich literature on single-cell downlink transmission that investigate the interplay between the feedback load and broadcast sum-rate throughput scaling. One group of studies, c.f. [8, 9, 10], for single-cell systems consider feeding back *unquantized* real values (channel magnitude or SNR, which in theory requires infinite-rate feedback. In another line of studies, c.f., [11], finite-rate feedback mechanisms are introduced. However, the optimal sum-rate throughput is achievable in these studies when the number of feedback bits tends to infinity. The proposed feedback mechanisms in [12, 13], impose only finite-rate feedback loads and achieve the optimal multiuser diversity gain. However, since they do not allow for simultaneous transmissions to multiple users, they cannot capture the multiplexing gain. Finally, the heuristic feedback mechanisms proposed in [14, 15] have limited feedback loads for which the sum-rate throughput scaling analysis is not provided. As a by-product of our analysis we find that in a *single*-cell downlink transmission it is possible to achieve the optimal capacity scaling rate with *finite*-rate feedback. It is also noteworthy on a different research direction, that capacity scaling for some infrastructure-based sensor networks are provided in [16], [17]

2 System Model

Consider the downlink transmission in a MIMO network that consists of M super-cells, each comprising of Q regular cells equipped with one base-station. We use the shorthand $\mathcal{Q}_m$ to refer to the mth super-cell and denote its base-stations by B_m^r, for $r \in \{1, \ldots, Q\}$. Let K denote the number of users per super-cell and assume that the base-stations of each super-cell accomplish the downlink transmissions collaboratively. Moreover, we assume that all the base-stations within a super-cell are fully cooperative, i.e., are connected through a backbone link that allows them to share their information streams and to carry out fully coordinated downlink transmissions. The case that such cooperation is not viable corresponds to the setting $Q = 1$. We also use the shorthand U_n^k to denote the kth user in the nth super-cell, respectively, for $k \in \{1, \ldots, K\}$ and $n \in \{1, \ldots, M\}$.

Define $\boldsymbol{H}_{n,m}^{k,r} \in \mathbb{C}^{N_r \times N_t}$ as the downlink channel from base-station B_m^r to user U_n^k and let $\boldsymbol{x}_m^r(t) \in \mathbb{C}^{N_t \times 1}$ denote the signal vector transmitted by base-station B_m^r at time instance t. The

received signal by U_n^k at time instance t is therefore given by

$$\boldsymbol{y}_n^k(t) = \sum_{m=1}^{M} \sum_{r=1}^{Q} \sqrt{\gamma_{n,m}^{k,r}} \; \boldsymbol{H}_{n,m}^{k,r} \boldsymbol{x}_m^r(t) + \boldsymbol{z}_n^k(t) \; ,$$

where $\gamma_{n,m}^{k,r} \in \mathbb{R}^+$ is incorporated to capitalize the combined effect of transmission power and signal attenuation (path-loss) between B_m^r and U_n^k. In non-homogenous networks the terms $\{\gamma_{n,m}^{k,r}\}$ are distinct as different users undergo different path-loss and shadowing effects. Also, $\boldsymbol{z}_n^k \in \mathbb{C}^{N_r \times 1}$ accounts for the additive white Gaussian noise distributed as $\mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_r})$. We consider block frequency-flat fading channels and assume that the entries of the channel matrix $\boldsymbol{H}_{n,m}^{k,r}$ are independent and identically distributed as complex Gaussian $\mathcal{CN}(0,1)$. In order to facilitate spatial multiplexing and to harness its ensuing interference as well as the inter-cell interference, the base-stations employ linear precoding. We define $\{\boldsymbol{\omega}_m^r(\ell)\}_{\ell=1}^{N_t}$ as the set of linear precoders employed by base-station B_m^r. By further defining $x_m^r(\ell, t)$ as the information symbol that B_m^r transmits along the beamforming vector $\boldsymbol{\omega}_m^r(\ell)$, at any time instance t we have

$$\boldsymbol{x}_m^r(t) = \sum_{\ell=1}^{N_t} x_m^r(\ell, t) \, \boldsymbol{\omega}_m^r(\ell) \; . \tag{1}$$

Motivated by refraining from information feedback about channel directions the base-stations exploit the notion of orthogonal random beamforming first proposed in [8]. For this purpose, at the beginning of each block transmission, each base-station generates N_t orthonormal vectors $\{\boldsymbol{\omega}_m^r(\ell)\}_{\ell=1}^{N_t}$, according to an isotropic distribution [18]. The user U_n^k knows $\boldsymbol{H}_{n,n}^{k,r}\boldsymbol{\omega}_n^r(\ell)$ for $\ell \in \{1, \ldots, N_t\}$ and $r \in \{1, \ldots, Q\}$ perfectly and instantaneously. We denote the transmission power per information stream by $\rho \triangleq \frac{P}{N_t}$, where P is the average power per base-station. Also, throughout the paper, we use the convention $a_K \doteq b_K$ to denote asymptotic equality, i.e., $\lim_{K \to \infty} \frac{a_K}{b_K} = 1$. Operators $\dot{\leq}$ and $\dot{\geq}$ are defined accordingly. All the logarithms, unless otherwise mentioned, are in base 2 and throughput are in bits/sec/Hz.

3   Scheduling Framework

We first consider the case that the users have one receive antenna and will later generalize it the case of arbitrary number of receive antennas in Section 4. We assume that the users employ single-user decoders such that all the inter-cell and intra-cell interferers are treated as Gaussian noise. While more advanced receivers will offer better performance, we show that the simple single-user

decoders are sufficient to fully capture the optimal multiplexing and multiuser diversity gains in MIMO networks.

Due to full cooperation, the information stream that some base-station in the super-cell $\mathcal{Q}_n$ intends to transmit to user U_n^k is known to all base-stations of that super-cell. Therefore, the message intended for user U_n^k can be transmitted along any arbitrary beamforming vector of any base-station of $\mathcal{Q}_n$. Hence, user U_n^k can compute the following $Q \times N_t$ distinct SINRs given that its intended information is embedded in $x_n^r(\ell, t)$ and transmitted by the base-station B_m^r along $\boldsymbol{\omega}_n^r(\ell)$, where $r \in \{1, \ldots, Q\}$ and $\ell \in \{1, \ldots, N_t\}$,

$$\mathsf{SINR}_n^k(r, \ell) = \frac{\gamma_{n,n}^{k,r} |\boldsymbol{H}_{n,n}^{k,r} \boldsymbol{\omega}_n^r(\ell)|^2}{\sum_{m=1}^{M} \sum_{r'=1}^{Q} \gamma_{n,m}^{k,r'} \sum_{\ell'=1}^{N_t} |\boldsymbol{H}_{n,m}^{k,r'} \boldsymbol{\omega}_m^{r'}(\ell')|^2 - \gamma_{n,n}^{k,r} |\boldsymbol{H}_{n,n}^{k,r} \boldsymbol{\omega}_n^r(\ell)|^2 + \frac{1}{\rho}} \ . \tag{2}$$

It can be readily verified that for any given ℓ and for any $(k, n, r) \neq (k', n', r')$, the numerators and denumerators of $\mathsf{SINR}_n^k(r, \ell)$ and $\mathsf{SINR}_{n'}^{k'}(r', \ell)$ are disjoint. In other words, any term of the form $\boldsymbol{H}_{n,m}^{k,r} \boldsymbol{\omega}_n^r(\ell)$ that appears in the characterization of $\mathsf{SINR}_n^k(r, \ell)$ does not appear in that of $\mathsf{SINR}_{n'}^{k'}(r', \ell)$. By invoking the statistical independence of the channel coefficients and beamforming vectors, we find that $\mathsf{SINR}_n^k(r, \ell)$ and $\mathsf{SINR}_{n'}^{k'}(r', \ell)$ are also statistically independent for any $(k, n, r) \neq (k', n', r')$. However, by noting the different attenuation factors $\{\gamma_{n,m}^{k,r}\}$ for different choices of k, n, r, $\mathsf{SINR}_n^k(r, \ell)$ and $\mathsf{SINR}_{n'}^{k'}(r', \ell)$ are not identically distributed. Similarly it can be shown that for any fixed set of k, n, r the terms $\{\mathsf{SINR}_n^k(r, \ell)\}$ are statistically identical but not independent. We denote the cumulative distribution function (CDF) of $\mathsf{SINR}_n^k(r, \ell)$ for $\ell \in \{1, \ldots, N_t\}$ by

$$F_n^k(x; r) : \mathbb{R}_+ \to [0, 1] \ .$$

We propose a user scheduling procedure that aims at identifying and serving QN_t users in each super-cell such that 1) the network sum-rate throughput captures the full multiuser diversity gain, 2) does not impose any information exchange or cooperation among the super-cells, and 3) necessitates each base-station to acquire only a *finite* number of feedback bits from each of its designated users. The steps involved in the proposed scheduling procedure are summarized as follows. We remark that the notion of random beamforming, first deployed in [8], has been the core of many scheduling schemes for single-cell broadcast channels.

1) Normalizing SINRs: For each user U_n^k we define Q positive scalars $\{\beta_n^k(1), \ldots, \beta_n^k(Q)\}$ each pertaining to one of the base-station of the super-cell $\mathcal{Q}_n$.

$$\forall r \in \{1, \ldots, Q\} : \quad \text{define} \quad \beta_n^k(r) \text{ such that } F_n^k\big(\beta_n^k(r); r\big) = 1 - \frac{1}{K} \ . \tag{3}$$

There exist *unique* solutions for $\beta_n^k(r)$ as $F_n^k(x;r)$ is a CDF and strictly monotonic in x. Next, corresponding to each user U_n^k we define the following normalized SINR terms.

$$\text{sinr}_n^k(r,\ell) \triangleq \frac{\text{SINR}_n^k(r,\ell)}{\beta_n^k(r)} \ , \quad \forall r \in \{1,\dots,Q\} \quad \text{and} \quad \ell \in \{1,\dots,N_t\} \ . \tag{4}$$

2) Feedback:

1. Recall that each user U_n^k can be served along any beamformer of the super-cell $\mathcal{Q}_n$. Among all beamformers, we identify the most favorable beam for the user U_n^k, that is the beamformer yielding the highest SINR for U_n^k, i.e.,

$$\overline{\text{sinr}}_n^k \triangleq \max_{r,\,\ell} \text{sinr}_n^k(r,\ell) \ . \tag{5}$$

2. If $\overline{\text{sinr}}_n^k \geq 1$, then user U_n^k feeds back the index of its most favorable beam to the base-stations of the super-cell $\mathcal{Q}_n$. As a super-cell has QN_t beamformers, conveying such feedback requires $\lceil \log QN_t \rceil$ information bits.

3) User Selection:

1. Based on the information fed back by the users, corresponding to each super-cell $\mathcal{Q}_n$ we construct the sets $\mathcal{H}_n^r(\ell)$, for $r \in \{1,\dots,Q\}$ and $\ell \in \{1,\dots,N_t\}$, such that $\mathcal{H}_n^r(\ell)$ contains the indices of the users in $\mathcal{Q}_n$ who have declared $\boldsymbol{\omega}_n^r(\ell)$ as their most favorable beam. By invoking (5) we have

$$\mathcal{H}_n^r(\ell) \triangleq \left\{ k \mid \overline{\text{sinr}}_n^k \geq 1 \quad \text{and} \quad \overline{\text{sinr}}_n^k = \text{sinr}_n^k(r,\ell) \right\} \ . \tag{6}$$

2. Super-cell $\mathcal{Q}_n$ *randomly* selects one user from each set $\mathcal{H}_n^r(\ell)$ for $r \in \{1,\dots,Q\}$ and $\ell \in \{1,\dots,N_t\}$ and embeds the information signal intended for this user in the stream $x_n^r(\ell)$. Note that the sets $\{\mathcal{H}_n^r(\ell)\}_{r,\ell}$ are mutually disjoint and as a result, it is guaranteed that a user will not be served by more than one beam.

This procedure involves information exchange only in the form of feedback from the users to their designated base-stations and the super-cells do not cooperate or exchange any information. The crucial part of such scheduling is the appropriate design of the normalization factors $\beta_n^k(r)$ that results in a user selection yielding the optimal capacity scaling. The choices of the normalization

factors as characterized in (3) serve a two-fold purpose, namely they control the aggregate amount of feedback and maintain fairness. As the network size K increases, the normalization factors $\beta_n^k(r)$ as defined in (3) affect the feedback load in two ways. On one hand more users are competing for accessing the resources and it is expected that more users participate in the feedback process. On the other hand, as K increases the normalization factors increase, which makes it harder for individual users in a network to satisfy the conditions for feeding back information to their designated base stations. As shown later in the paper, the combined impact of these two effects is in favor of reducing the feedback load. Also, as for the fairness, incorporating the statistics of the SINRs in the choices of the normalization factors ensures that all the users are equiprobable in satisfying their feedback constraint and consequently by invoking the selection procedure, all are equiprobable in having access to the resources. The proposed scheduling mechanism requires the users acquire the SINRs yielded by different beamformers. Such SINR acquisition can be accomplished during the training phase when the users obtain the channel states prior to the downlink transmission of data.

4   Throughput Scaling Analysis

When in a MIMO network perfect cooperation (full information exchange) among all super-cells is allowed, it can be effectively modeled as a large single-cell multiple-antenna broadcast channel with MQN_t transmit antennas and MK users. In [3, 8] it shown that for such downlink transmission the sum-rate capacity, which is achievable via dirty-paper coding, scales with increasing K as

$$\mathbb{E}_{\boldsymbol{H}}[R] \doteq MQN_t \log \log KN_r \ . \tag{7}$$

In the sequel we analyze the sum-rate throughput scaling of our proposed procedure and demonstrate that the achievable throughput scales similar to (7).

4.1   Throughput Scaling for $N_r = 1$

By defining R_n as the sum-rate throughput achieved for the nth super-cell, the expected sum-rate throughput of the network is $\mathbb{E}_{\boldsymbol{H}}[R] = \sum_{n=1}^{M} \mathbb{E}_{\boldsymbol{H}}[R_n]$. We further define $R_n^r(\ell)$ as the rate supported through the transmission along the beamforming vector $\boldsymbol{\omega}_n^r(\ell)$. As formalized in (6), the super-cell $\mathcal{Q}_n$ constructs the new set $\bar{\mathcal{H}}_n$, as the set of the user index sets $\mathcal{H}_n^r(\ell)$ for all r and ℓ, i.e., $\bar{\mathcal{H}}_n \triangleq \{\mathcal{H}_n^r(\ell)\}_{r,\ell}$. Conditioned on the sets $\bar{\mathcal{H}}_n$, the sum-rate throughput supported in the nth

super-cell is the summation of the rates its individual beamformers can sustain. Therefore,

$$\mathbb{E}_{\boldsymbol{H}}[R_n \mid \bar{\mathcal{H}}_n] = \sum_{r=1}^{Q} \sum_{\ell=1}^{N_t} \mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] \ . \tag{8}$$

As designated by the user scheduling procedure, for each $r \in \{1, \ldots, Q\}$ and $\ell \in \{1, \ldots, N_t\}$, one of the members of the set $\mathcal{H}_n^r(\ell)$ is opted *randomly* to be served along the beamforming vector $\boldsymbol{\omega}_n^r(\ell)$. Hence, each user with its index in $\mathcal{H}_n^r(\ell)$ will be selected to be served along the beamformer $\boldsymbol{\omega}_n^r(\ell)$ with probability $\frac{1}{|\mathcal{H}_n^r(\ell)|}$. As a result given the set of candidate users $\mathcal{H}_n^r(\ell)$, the expected value of the rate sustained by the beamformer $\boldsymbol{\omega}_n^r(\ell)$ is equal to the arithmetic average of the rates these candidate users can sustain. By noting that if the transmission of the beamformer $\boldsymbol{\omega}_n^t(\ell)$ is intended for user U_n^k then $R_n^r(\ell) = \log(1 + \mathsf{SINR}_n^k(r, \ell))$, we obtain

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] = \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \mathcal{H}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + \mathsf{SINR}_n^k(r, \ell) \right) \mid \mathcal{H}_n^r(\ell) \right] . \tag{9}$$

The next lemma, which provides a lower bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ is instrumental to analyzing the scaling rate of $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$.

Lemma 1 (Conditional Expected Rates) *If $\beta_n^k(r) \geq 1 \ \forall n \in \{1, \ldots, M\}, \ \forall k \in \{1, \ldots, K\}$, and $\forall r \in \{1, \ldots, Q\}$, then*

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] \ \geq \ \mathcal{R}_n^r(\ell, |\mathcal{H}_n^r(\ell)|) \ \triangleq \ \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{j=1}^{|\mathcal{H}_n^r(\ell)|} \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + \mathsf{SINR}_n^{(j)}(r, \ell) \right) \right] , \tag{10}$$

where $\mathsf{SINR}_n^{(j)}(r, \ell)$ denotes the jth largest element of the set $\{\mathsf{SINR}_n^1(r, \ell), \ldots, \mathsf{SINR}_n^K(r, \ell)\}$.

Proof: See Appendix A. ∎

Equations (8) and (9) along with Lemma 1 (conditional expected rates) establish that if $\forall k, n, r$ the normalizing factors satisfy the constraints $\beta_n^k(r) \geq 1$, then we have

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{H}}[R_n] \ &= \ \sum_{\bar{\mathcal{H}}_n} P(\bar{\mathcal{H}}_n) \, \mathbb{E}_{\boldsymbol{H}}[R_n \mid \bar{\mathcal{H}}_n] \\
&\overset{(8)}{=} \ \sum_{\bar{\mathcal{H}}_n} P(\bar{\mathcal{H}}_n) \sum_{r=1}^{Q} \sum_{\ell=1}^{N_t} \mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] \\
&\overset{(10)}{\geq} \ \sum_{\bar{\mathcal{H}}_n} P(\bar{\mathcal{H}}_n) \sum_{r, \, \ell} \mathcal{R}_n^r(\ell, |\mathcal{H}_n^r(\ell)|) \ .
\end{aligned}
\tag{11}
$$

Note that the rates $\mathcal{R}_n^r(\ell, |\mathcal{H}_n^r(\ell)|)$ are identical for all permutations of the set $\mathcal{H}_n^r(\ell)$ that have the same cardinality. In other words, $\mathcal{R}_n^r(\ell, |\mathcal{H}_n^r(\ell)|)$ depends on $\mathcal{H}_n^r(\ell)$ only through its cardinality

which can be any member of $\{0, \ldots, K\}$. Note that (11) is a weighted sum of the rates $\mathcal{R}_n^r(\ell, b)$ for $\ell \in \{1, \ldots, N_t\}$, $r \in \{1, \ldots, Q\}$, and $b \in \{1, \ldots, K\}$, and in order to further simplify (11) we need to find their weighting coefficients. For this purpose we rearrange all the relevant terms that contribute to $\mathcal{R}_n^r(\ell, k)$ and find that

$$\text{coefficient of } \mathcal{R}_n^r(\ell, b) = \sum_{\{\bar{\mathcal{H}}_n \mid |\mathcal{H}_n^r(\ell)| = b\}} P(\bar{\mathcal{H}}_n) = P(|\mathcal{H}_n^r(\ell)| = b) . \tag{12}$$

Therefore, combining (11) and (12) yields that if $\forall k, n, r$ the constraints $\beta_n^k(r) \geq 1$ are satisfied, then

$$
\begin{aligned}
\mathbb{E}_{\boldsymbol{H}}[R_n] \quad &\geq \quad \sum_{r,\,\ell,\,b} P(|\mathcal{H}_n^r(\ell)| = b)\, \mathcal{R}_n^r(\ell, b) \\
&\stackrel{(10)}{=} \quad \sum_{r,\,\ell,\,b} \frac{P(|\mathcal{H}_n^r(\ell)| = b)}{b} \sum_{j=1}^{b} \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right)\right] \\
&= \quad \sum_{r,\,\ell,\,j} \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right)\right] \underbrace{\sum_{b=j}^{K} \frac{P(|\mathcal{H}_n^r(\ell)| = k)}{k}}_{\triangleq\, q_n^j(r,\ell)} \tag{13} \\
&= \quad \sum_{r,\,\ell,\,j} q_n^j(r, \ell)\, \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right)\right] \triangleq \bar{R}_n . \tag{14}
\end{aligned}
$$

By setting $q_n^0(r, \ell) \triangleq P(|\mathcal{H}_n^r(\ell)| = 0)$ we find that for any fixed n, r, ℓ the sequence $\{q_n^1(r, \ell), \ldots, q_n^K(r, \ell)\}$ constitutes a probability mass function (PMF) as

$$
\begin{aligned}
\sum_{j=0}^{K} q_n^j(r, \ell) &= q_n^0(r, \ell) + \sum_{j=1}^{K} \sum_{k=j}^{K} \frac{P(|\mathcal{H}_n^r(\ell)| = k)}{k} \\
&= q_n^0(r, \ell) + \sum_{k=1}^{K} \frac{1}{k} \sum_{j=1}^{k} P(|\mathcal{H}_n^r(\ell)| = k) = \sum_{k=0}^{K} P(|\mathcal{H}_n^r(\ell)| = k) = 1 .
\end{aligned}
$$

Now, we assess the scaling behavior of $\bar{R}_n$ which in turn provides a lower bound on the throughput scaling of our proposed scheduling procedure. For this purpose, based on (14), we need to obtain the distributions of order statistics $\{\mathsf{SINR}_n^{(1)}(r, \ell), \ldots, \mathsf{SINR}_n^{(K)}(r, \ell)\}$ for $n \in \{1, \ldots, M\}$, $r \in \{1, \ldots, Q\}$, and $\ell \in \{1, \ldots, N_t\}$. As mentioned earlier the SINR terms $\{\mathsf{SINR}_n^1(r, \ell), \ldots, \mathsf{SINR}_n^K(r, \ell)\}$ are independent but *not* identical and thereof characterizing the distributions of their associated sequence of order statistics does not seem tractable. For tractability purposes, for each $n \in \{1, \ldots, M\}$, $r \in \{1, \ldots, Q\}$, and $\ell \in \{1, \ldots, N_t\}$ we construct two other sets one containing lower bounds and

the other containing upper bounds on $\{\mathsf{SINR}_n^1(r,\ell),\ldots,\mathsf{SINR}_n^K(r,\ell)\}$. We define

$$\zeta_n^1 = \min_{m,k,r} \gamma_{n,m}^{k,r} \qquad \text{and} \qquad \eta_n^1 = \min_{k,r} \gamma_{n,n}^{k,r} ,$$

$$\zeta_n^2 = \max_{m,k,r} \gamma_{n,m}^{k,r} \qquad \text{and} \qquad \eta_n^2 = \max_{k,r} \gamma_{n,n}^{k,r} .$$

Based on these we define the following lower bound on $\mathsf{SINR}_n^k(r,\ell)$

$$S_n^k(r,\ell) \triangleq \frac{|\boldsymbol{H}_{n,n}^{k,r}\boldsymbol{\omega}_n^r(\ell)|^2}{\frac{\zeta_n^2}{\eta_n^1}\left(\sum_{m,r',\ell'}|\boldsymbol{H}_{n,m}^{k,r'}\boldsymbol{\omega}_n^{r'}(\ell')|^2 - |\boldsymbol{H}_{n,n}^{k,r}\boldsymbol{\omega}_n^r(\ell)|^2\right) + \frac{1}{\rho\eta_n^1}} , \tag{15}$$

and the following upper bound on $\mathsf{SINR}_n^k(\ell)$

$$T_n^k(r,\ell) \triangleq \frac{|\boldsymbol{H}_{n,n}^{k,r}\boldsymbol{\omega}_n^r(\ell)|^2}{\frac{\zeta_n^1}{\eta_n^2}\left(\sum_{m,r',\ell'}|\boldsymbol{H}_{n,m}^{k,r'}\boldsymbol{\omega}_n^{r'}(\ell')|^2 - |\boldsymbol{H}_{n,n}^{k,r}\boldsymbol{\omega}_n^r(\ell)|^2\right) + \frac{1}{\rho\eta_n^2}} , \tag{16}$$

Next we show that $S_n^k(r,\ell)$ and $T_n^k(r,\ell)$ and their associated order statistics have tractable distributions.

Lemma 2 (SINR Bounds Distributions) *For any given $n \in \{1,\ldots,M\}$, $r \in \{1,\ldots,Q\}$, and $\ell \in \{1,\ldots,N_t\}$, the elements of the set $\{S_n^k(r,\ell)\}_{k=1}^K$ are i.i.d. with CDF*

$$S_n^k(r,\ell) \sim F_1(x;n) \triangleq 1 - \frac{e^{-x/\rho\eta_n^1}}{\left(\frac{\zeta_n^2}{\eta_n^1}x+1\right)^{MQN_t-1}} , \tag{17}$$

and the elements of the set $\{T_n^k(r,\ell)\}_{k=1}^K$ are i.i.d. with CDF

$$T_n^k(r,\ell) \sim F_2(x;n) \triangleq 1 - \frac{e^{-x/\rho\eta_n^2}}{\left(\frac{\zeta_n^1}{\eta_n^2}x+1\right)^{MQN_t-1}} . \tag{18}$$

Proof: See Appendix B. ∎

Assessing the scaling of $\bar{R}_n$ defined in (14) (which is a lower bound for $\mathbb{E}_{\boldsymbol{H}}[R_n]$) requires the knowledge of the distributions of the order statistics $\{\mathsf{SINR}_n^{(1)}(r,\ell),\ldots,\mathsf{SINR}_n^{(K)}(r,\ell)\}$. In the next lemma it is shown how $\mathsf{SINR}_n^{(j)}(r,\ell)$ is related to the order statistics of the sets $\{S_n^k(r,\ell)\}_{k=1}^K$ and $\{T_n^k(r,\ell)\}_{k=1}^K$. We denote the jth largest elements of the sets $\{S_n^k(r,\ell)\}_{k=1}^K$ and $\{T_n^k(r,\ell)\}_{k=1}^K$ by $S_n^{(j)}(r,\ell)$ and $T_n^{(j)}(r,\ell)$, respectively.

Lemma 3 (Ordered SINR Bounds) *For any $n \in \{1,\ldots,M\}$, $r \in \{1,\ldots,Q\}$, and $\ell \in \{1,\ldots,N_t\}$, we have*

$$\forall j \in \{1,\ldots,K\}: \quad S_n^{(j)}(r,\ell) \leq \mathsf{SINR}_n^{(j)}(r,\ell) \leq T_n^{(j)}(r,\ell) .$$

10

Proof: See Appendix C. ■

By invoking Lemma 3 (ordered SINR bounds), from (14) we find that if $\forall k, n, \beta(n, k) \geq 1$, then we have following bounds on $\bar{R}_n$.

$$\sum_{r, \ell, j} q_n^j(r, \ell) \, \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + S_n^{(j)}(r, \ell) \right) \right] \leq \bar{R}_n \leq \sum_{r, \ell, j} q_n^j(r, \ell) \, \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + T_n^{(j)}(r, \ell) \right) \right]. \quad (19)$$

By denoting the CDFs of $S_n^{(j)}(r, \ell)$ and $T_n^{(j)}(r, \ell)$ by $F_1^{(j)}(x; n)$ and $F_2^{(j)}(x; n)$, respectively, (19) can be restated as

$$\sum_{r, \ell} \int_0^\infty \log(1 + x) \, d\left(\sum_{j=1}^K q_n^j(r, \ell) F_1^{(j)}(x; n) \right) \leq \bar{R}_n \leq \sum_{r, \ell} \int_0^\infty \log(1 + x) \, d\left(\sum_{j=1}^K q_n^j(r, \ell) F_2^{(j)}(x; n) \right). \quad (20)$$

Note that

$$\{q_n^j(r, \ell)\}_j \text{ is a PMF} \quad \Rightarrow \quad \sum_{j=1}^K q_n^j(r, \ell) \, F_1^{(j)}(x; n) \text{ and } \sum_{j=1}^K q_n^j(r, \ell) \, F_2^{(j)}(x; n) \text{ are CDFs },$$

and thereof we define the new CDFs

$$F_1^K(x; n, r, \ell) \triangleq \sum_{j=1}^K q_n^j(r, \ell) \, F_1^{(j)}(x; n) \quad \text{and} \quad F_2^K(x; n, r, \ell) \triangleq \sum_{j=1}^K q_n^j(r, \ell) \, F_2^{(j)}(x; n) .$$

Hence, we can further reformat (20) as

$$\sum_{r, \ell} \int_0^\infty \log(1 + x) \, dF_1^K(x; n, r, \ell) \leq \bar{R}_n \leq \sum_{r, \ell} \int_0^\infty \log(1 + x) \, dF_2^K(x; n, r, \ell) . \quad (21)$$

In the next step, we define an auxiliary random variable X distributed exponentially with unit mean and denote its CDF by

$$G(x) \triangleq 1 - e^{-x} .$$

Also let $G^{(j)}(x)$ denote the CDF of the jth order statistic of a set of statistical samples with K members with the parent distribution $G(x)$. Similar to $F_1^K(x; n, \ell)$ and $F_2^K(x; n, \ell)$ define

$$G^K(x; n, r, \ell) \triangleq \sum_{j=1}^K q_n^j(r, \ell) \, G^{(j)}(x) .$$

The motivation for defining this auxiliary exponential random variable is to ultimately obtain lower and upper bounds on $\bar{R}_n$ that can be characterized as functions of $G(x)$. Such characterization is instrumental in the sense that we will be able to use some well-known properties of the order statistics of exponential distribution for finding the scaling behavior of $\bar{R}_n$. The next lemma provides the key step for connecting the bounds given in (21) with new bounds that are functions of $G(x)$.

Lemma 4 (Bounds on $\bar{R}_n$) *If $\beta_n^k(r) \geq 1 \; \forall n \in \{1, \ldots, M\}, \; \forall k \in \{1, \ldots, K\}, \; and \; \forall r \in \{1, \ldots, Q\},$ then we have*

$$\sum_{r, \ell} \int_0^\infty \log(1 + \rho \eta_n^1 x) \, dG^K(x; n, r, \ell) - A_n \leq \bar{R}_n \leq \sum_{r, \ell} \int_0^\infty \log(1 + \rho \eta_n^2 x) \, dG^K(x; n, r, \ell) \; , \quad (22)$$

where $A_n \triangleq QN_t \log\left((MQN_t - 1)\rho\zeta_n^2\eta_n^1 + 1\right).$

Proof: See Appendix D. ∎

To this end we have established lower and upper bounds on $\bar{R}_n$ based on the order statistics of the exponential distribution. As the final step, we assess the scaling of these bounds. The next lemma reveals how the scaling of the bounds on $\bar{R}_n$ provided in Lemma 4 (bounds on $\bar{R}_n$), pertains to the characterization of the CDF $G^K(x; \ell)$, which itself depend on the PMFs $\{q_n^j(r, \ell)\}_j$.

Lemma 5 (Scaling) *For a sequence of K i.i.d unit-variance exponentially distributed random variables with CDF $W(x)$ and for any arbitrary PMF $\{Q_j\}_{j=1}^K$ if the condition*

$$\lim_{K \to \infty} \frac{\sum_{j=1}^K jQ_j}{K} = 0 \; , \quad (23)$$

is satisfied, then for any positive real number $a \in \mathbb{R}_+$ we have

$$\int_0^\infty \log(1 + ax) dW^K(x) \doteq \log\log K + \log a \; ,$$

where

$$W^K(x) \triangleq \sum_{j=1}^K Q_j W^{(j)}(x) \; ,$$

and $W^{(j)}$ is the CDF of the jth order statistic of the set of these K random variables.

Proof: See Appendix E. ∎

By using the results of Lemmas 4 (bounds on $\bar{R}_n$) and 5 (scaling) we state the main result of this paper in the next theorem.

Theorem 1 (MISO Capacity Scaling) *For downlink transmission in a MISO network with M super-cells each with Q base-stations, K users per super-cell and N_t transmit antenna per base-station, the sum-rate throughput achievable via the proposed scheduling procedure scales as*

$$\mathbb{E}_{\boldsymbol{H}}[R] \doteq MQN_t \log\log K \; .$$

Proof: We start by showing that the choice of $Q_j = q_n^j(r, \ell)$ for $j \in \{1, \ldots, K\}$, where $\{q_n^j(r, \ell)\}$ is defined in (11), fulfils the condition (13) of Lemma 5 (scaling) for any $n \in \{1, \ldots, M\}$, $r \in \{1, \ldots, Q\}$, and $\ell \in \{1, \ldots, N_t\}$. From (13) we have

$$
\begin{aligned}
\sum_{j=1}^{K} j Q_j &= \sum_{j=1}^{K} j \, q_n^j(r, \ell) = \sum_{j=1}^{K} j \sum_{k=j}^{K} \frac{P(|\mathcal{H}_n^r(\ell)| = k)}{k} \\
&= \sum_{k=1}^{K} \frac{1}{k} P(|\mathcal{H}_n^r(\ell)| = k) \sum_{j=1}^{k} j \\
&= \sum_{k=1}^{K} \frac{k+1}{2} P(|\mathcal{H}_n^r(\ell)| = k) \\
&= \frac{1}{2} \underbrace{\sum_{k=1}^{K} k \, P(|\mathcal{H}_n^r(\ell)| = k)}_{\mathbb{E}[|\mathcal{H}_n^r(\ell)|]} + \frac{1}{2} \underbrace{\sum_{k=0}^{K} P(|\mathcal{H}_n^r(\ell)| = k)}_{=1} - \frac{1}{2} P(|\mathcal{H}_n^r(\ell)| = 0) \\
&\leq \frac{1}{2} \mathbb{E}[|\mathcal{H}_n^r(\ell)|] + \frac{1}{2} \, .
\end{aligned}
\tag{24}
$$

Note that the probability that each individual user in the nth super-cell expresses interest in the ℓth beam of B_n^r, i.e., $\boldsymbol{\omega}_n^r(\ell)$, by feeding back its index is $\frac{1}{K}$. As shown at the beginning of Appendix A we have

$$
\text{if } \beta_n^k(r) \geq 1 \quad \Rightarrow \quad \left\{ \quad \mathsf{SINR}_n^k(r, \ell) \geq \beta_n^k(r) \quad \Leftrightarrow \quad k \in \mathcal{H}_n^r(\ell) \quad \right\} \, .
$$

Therefore,

$$
P(k \in \mathcal{H}_n^r(\ell)) = P\left(\mathsf{SINR}_n^k(r, \ell) \geq \beta_n^k(r)\right) = 1 - F_n^k(\mathsf{SINR}_n^k(r, \ell); r) \overset{(3)}{=} \frac{1}{K} \, .
$$

Since $|\mathcal{H}_n^r(\ell)|$ counts the number of users in the nth super-cell that have expressed interest in beamformer $\boldsymbol{\omega}_n^r(\ell)$, we have

$$
|\mathcal{H}_n^r(\ell)| \sim \text{Binomial}(K, 1/K) \quad \Rightarrow \quad \mathbb{E}[|\mathcal{H}_n^r(\ell)|] = K \cdot \frac{1}{K} = 1 \, .
\tag{25}
$$

Equations (24) and (25) provide that

$$
\sum_j j Q_j \leq 1 \quad \Rightarrow \quad \lim_{K \to \infty} \frac{\sum_{j=1}^{K} j Q_j}{K} = 0 \, .
$$

Therefore, the necessary condition of Lemma 5 (scaling) given in (23) is satisfied and we can apply Lemma 5 (scaling) on the bounds on $\bar{R}_n$ provided in Lemma 4 (bounds on $\bar{R}_n$). By setting

$$
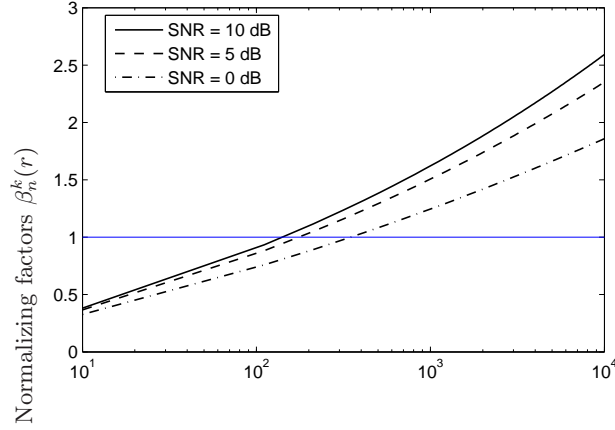W(x) = G(x) \quad , \forall j \; Q_j = q_n^j(r, \ell), \quad \text{and} \quad a = \rho \eta_n^1, \rho \eta_n^2 \, ,
$$

Figure 1: Normalizing factors $\beta_n^k(r)$ versus the number of users per super-cell K, for $M = 3$, $Q = 2$ $N_t = 2$ and $\gamma_{n,m}^{k,r} = 1$.

according to Lemma 4 (bounds on $\bar{R}_n$) we find

$$QN_t \log \log K + QN_t \log \left(\frac{\rho \eta_n^1}{(MQN_t - 1)\rho \zeta_n^2 \eta_n^1 + 1} \right) \le \bar{R}_n \le QN_t \log \log K + QN_t \log(\rho \eta_n^2) \ .$$

The bounds above demonstrate that for fixed M, N_t, ρ, η_n^1 and η_n^2, and for increasing K we have

$$\lim_{K \to \infty} \frac{\bar{R}_n}{QN_t \log \log K} = 1 \quad \Rightarrow \quad \bar{R}_n \doteq QN_t \log \log K \ .$$

Taking into account that the sum-rate throughput cannot scale faster than what is achieved via dirty paper coding, i.e., $MQN_t \log \log K$, the proof is complete. ∎

All the analyses are constructed on the basis that $\forall k, n, r$, we have $\beta_n^k(r) \ge 1$. As discussed earlier, under this condition, each user will have only one of its normalized SINRs (corresponding to different beamformers) greater than 1. Also, as implied throughout the analysis, at the asymptote of large K, in each super-cell only QN_t users will participate in the feedback processes, which is the core reason of having controlled amount of feedback. The condition $\beta_n^k(r) \ge 1$ can be guaranteed in the asymptote of large K by recalling that $F_n^k(\beta_n^k(r); r) = 1 - \frac{1}{K}$. By numerically evaluating $F_n^k(\beta_n^k(r); r)$ through Monte Carlo realizations, Fig. 1 shows how the normalization factor changes with the number of users at SNRs. It is seen that in a network with $M = 3$ super-cells, each containing $Q = 2$ cells and each base-station equipped with $N_t = 2$ transmit antennas the normalizing factors $\beta_n^k(r)$ very quickly exceed 1 (as the requirement of Lemma 1 on the conditional expected rates) for different SNR levels $0, 5$ and 10 dB. The normalization factors are obtained via numerically solving (3).

4.2 Throughput Scaling for $N_r > 1$

In this section we generalize the result of Theorem 1 (MISO capacity scaling) to MIMO networks with arbitrary number of receive antennas. As stated earlier in (7), by employing dirty-paper coding and facilitating full CSI feedback, the sum-rate throughput exhibits a double logarithmic growth with the number of receive antennas, N_r. We show that with slight modification to the scheduling procedure provided in Section 3 the same gain can be retained.

We modify the scheduling procedure such that it allows each user to receive more than one information stream across its different receive antennas. In other words, we preclude the receive antennas to jointly decode the information streams and consider different receive antennas as separate users. As a result, this translates to having effectively KN_r users in the network and all the analyses provided earlier for K users can be extended for the network with KN_r users. It is noteworthy that although it might not be the most effective way to exploit the receive antennas, it yet achieves the optimal capacity scaling. Indeed, more sophisticated approaches involving joint signal decoding by the antenna arrays are expected to achieve the same gain. Based on this discussion, we modify the normalization factors $\beta_n^k(r)$ such that they satisfy

$$F_n^k(\beta_n^k(r); r) = 1 - \frac{1}{KN_r} \ , \tag{26}$$

and use the same normalization factor for all antennas of the same user. According to the modified scheduling algorithm on one hand, each user might be required to feed back the indices of more than one beam, which potentially increases the amount of feedback amount per user. On the other hand, increasing the normalization factor makes it more stringent for the users to satisfy their designated normalization factor constraints and send feedback. The discussions on the aggregate amount of feedback and effects of the number of users and number of receive antennas are given in Section 5. As a direct result of Theorem 1 (MISO capacity scaling) we obtain the following scaling law for MIMO networks.

Theorem 2 (MIMO Capacity Scaling) *For downlink transmission in a MIMO network with M super-cells each with Q base-stations, K users per super-cell, N_t transmit antenna per base-station, and N_r receive antennas pre user, the sum-rate throughput achievable via the proposed scheduling procedure scales as*

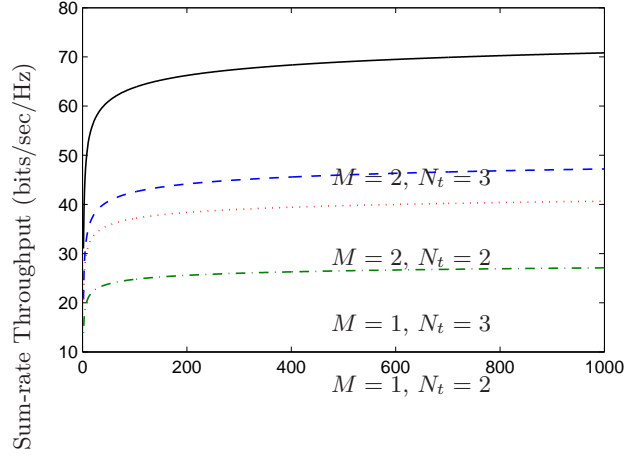$$\mathbb{E}_{\boldsymbol{H}}[R] \doteq MQN_t \log \log KN_r \ .$$

Figure 2: Sum-rate throughput versus the number of users per super-cell K, for $Q = 2$ and $\rho = 10$ dB. Each user has $N_r = 1$ receive antenna.

We remark that in the scheduling procedure we have considered that each user may receive more than one information stream. On the other hand, if we restrict each user to receive no more than one information stream, the performance will degrade. Indeed, as shown in [8] for single-cell broadcast transmissions, even with perfect SINR feedback, the sum-rate throughput scales as $N_t \log \log K$ which does not depend on the number of receive antennas. Therefore, with such restriction, adding more receiver antennas offers no asymptotic gain. Simulation results in Fig. 2 depict the sum-rate throughput achieved by the proposed scheduling scheme and show that it varies linearly with the number of super-cells M and transmit antennas N_t and double logarithmically with the number of users K.

Finally, as a corollary to Theorem 2 we can recover the existing results on the capacity scaling of the *single*-cell broadcast channels. The result provided in the following corollary indicates that deploying the proposed scheduling scheme for only one super-cell, requires only *finite*-rate feedback and yet retains the optimal capacity scaling of the broadcast channels with full channel state information feedback. Similar to the results for the multi-cell networks, in the following corollary we also assume that the number of transmit and receive antennas N_t, N_r are fixed.

Corollary 1 (Single-cell Capacity Scaling) *For a MIMO broadcast channel with N_t transmit antenna that serves K users each equipped with N_r receive antennas, the sum-rate throughput achievable via the proposed scheduling procedure scales as*

$$\mathbb{E}_{\boldsymbol{H}}[R] \doteq N_t \log \log K N_r \ .$$

16

5 Aggregate Feedback

So far we have established the sum-rate throughput scaling. In this section we evaluate the expected amount of feedback bits imposed by the proposed scheduling procedure. This essentially provides an upper bound on the minimum required amount of information exchange in the network. As discussed in Section 3, if a user meets the criterion to be considered as a candidate for being selected and scheduled, it will feed back $\lceil \log QN_t \rceil$ bits per eligible receive antenna, and otherwise will remain silent indicating that it is not a candidate for being scheduled. Therefore, each user, potentially, might send as many as $\min(N_t, N_r)\lceil \log QN_t \rceil$ bits via the feedback channel. This, in turn, can potentially give rise to up to $K \min(N_t, N_r)\lceil \log QN_t \rceil$ aggregate feedback bits in each super-cell, which can be prohibitive. Therefore, it is imperative to examine and control the burden put on the feedback channel.

Note that the normalization factors $\beta_n^k(r)$ change adaptively with varying K. Therefore, as the number of users per super-cell increases, the chance that each individual user satisfies the normalization constraint becomes slimmer. In other words, the choices of $\beta_n^k(r)$ guarantee that as the network becomes dense, the aggregate feedback load is controlled via enforcing a more stringent constraint on each user to utilize the feedback channel. In practice, obtaining $\beta_n^k(r)$ through solving (3) can be carried out via bi-sectional search. As the CDF is an strictly increasing function in its argument and the problem in (3) has a unique solution, a simple bi-section search, which is computationally efficient, can solve the problem with arbitrary accuracy. In the next theorem we quantify the aggregate feedback and show that the *expected* aggregate feedback is a constant independent of the number of users in the network.

Theorem 3 (Aggregate Feedback) *For downlink transmission in a MIMO network with M super-cells each with Q base-stations, K users per super-cell, N_t transmit antenna per base-station, and N_r receive antennas pre user, by deploying the proposed scheduling procedure, the aggregate feedback in the network is $MQN_t \lceil \log QN_t \rceil$.*

Proof: Lets define $N = \min(N_t, N_r)$, which is the maximum number of information streams a user can receive. Also let $\mathsf{SINR}_n^k(r, \ell, i)$ denote the detection SINR of ith receive antenna of user U_n^k when served along the beamforming vector $\boldsymbol{\omega}_n^r(\ell)$. The probability that the ith receive antenna of user U_n^k is a candidate to be served by one of the QN_t beamformers of its designated super-cell, is

upper bounded by

$$P\Big(\max_{r,\ell}\frac{\mathsf{SINR}_n^k(r,\ell,i)}{\beta_n^k(r)}\geq 1\Big)\overset{\text{union bound}}{\leq}\sum_{r=1}^{Q}\sum_{\ell=1}^{N_t}P\Big(\mathsf{SINR}_n^k(r,\ell,i)\geq\beta_n^k(r)\Big)=\frac{QN_t}{KN_r}\triangleq\mu\ .$$

Hence, the probability that exactly $j\leq N$ users satisfy the normalization constraint is upper bounded by $\mu_j\triangleq\binom{N_r}{j}\mu^j(1-\mu)^{N_r-j}$. Therefore, by denoting the number of feedback bits per user by R_n^{FB} and taking into account that if a user satisfies the normalization constraint, it will feed back $\lceil\log QN_t\rceil$ bits, we get

$$\mathbb{E}[R^{\mathrm{FB}}]\leq\lceil\log QN_t\rceil\sum_{j=1}^{N}j\mu_j=\lceil\log QN_t\rceil\sum_{j=1}^{N}j\binom{N_r}{j}\mu^j(1-\mu)^{N_r-j}.$$

Hence, the aggregate amount of feedback in the nth super-cell, denoted by $\mathbb{E}[R_n^{\mathrm{FB}}]$ in the asymptote of large K is

$$
\begin{aligned}
\mathbb{E}[R_n^{\mathrm{FB}}] &\leq \lim_{K\to\infty}K\lceil\log QN_t\rceil\sum_{j=1}^{N}j\binom{N_r}{j}\mu^j(1-\mu)^{N_r-j}\\
&= \lceil\log QN_t\rceil\sum_{j=1}^{N}j\binom{N_r}{j}\lim_{K\to\infty}\frac{\left(\frac{QN_t}{KN_r}\right)^j\left(1-\frac{QN_t}{KN_r}\right)^{N_r-j}}{\frac{1}{K}}\\
&= \lceil\log QN_t\rceil\sum_{j=1}^{N}j\binom{N_r}{j}\left(\frac{QN_t}{N_r}\right)^j\lim_{K\to\infty}\frac{1}{K^{j-1}}\\
&= QN_t\lceil\log QN_t\rceil\ ,
\end{aligned}
\tag{27}
$$

which provides that the expected aggregate amount of feedback bits in the network does not exceed $MQN_t\lceil\log QN_t\rceil$. The last equality holds by noting that for $j>1$ the term $\frac{1}{K^{j-1}}\to 0$ and the only non-zero term corresponds to $j=1$ and is equal to $QN_t\lceil\log QN_t\rceil$. ∎

The simulation results provided in Fig. 3 quantify the aggregate amount of feedback imposed by our scheduling protocol to the feedback channel. As anticipated by the analyses, as the number of users increase, the total amount of feedback bits is $QN_t\lceil\log QN_t\rceil$ which is independent of the number of users as well as the number of receive antennas. This implies that as the number of users increase, only QN_t users will participate in feeding back their indices of their desired beams. This is due to the fact that high normalizing factors for networks with large number of users, only allow the best QN_t users to feed back the indices of their corresponding beams. For the simulations we consider the settings $Q=1,2$ base-stations per super-cell and assume that each user has $N_r=3$ receive antennas.
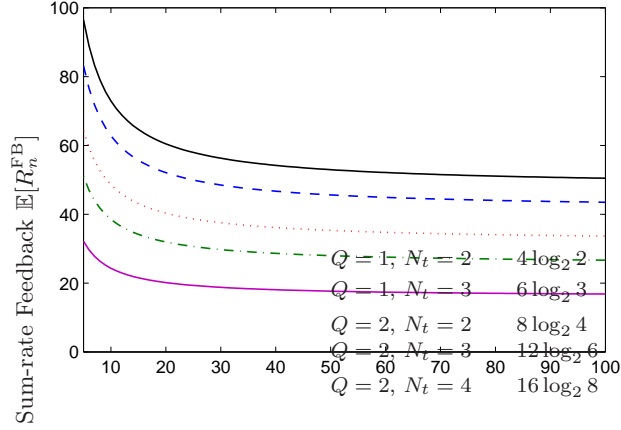
Figure 3: Aggregate amount of feedback versus the number of users per super-cell K. Each user has $N_r = 1$ receive antennas and $\rho = 10$ dB.

6 Practical Issues

6.1 Size of Super-cells

The size of the super-cells Q impacts both the scaling rate of the sum-rate throughput (Theorem 1 on the MISO capacity scaling) as well as the amount of information exchange in the network (Theorem 3 on aggregate feedback). The capacity scaling analysis shows that the sum-rate throughput scales as $MQN_t \log \log KN_r$. This is the scaling of a network with a total number of MQ base-stations and MK users, which means there exist $T \triangleq \frac{K}{Q}$ users per base-station. When keeping this ratio of users per base-station constant, the sum-rate throughput scales as $MQN_t \log \log QTN_r$. This scaling essentially implies that as long as the product MQ is a constant, the sum-rate scaling remains unchanged. The reason is that for fixed MQ, necessarily M will be a finite integer, based on which we have

$$\lim_{K \to \infty} \frac{\mathbb{E}_{\boldsymbol{H}}[R]}{MQN_t \log \log MKN_r} = \lim_{K \to \infty} \frac{\log \log KN_r}{\log \log MKN_r} = 1,$$

which implies that the throughput of the proposed procedure scales identical to the scaling law provided in Theorem 2. Therefore, when the total number of base-stations in the network MQ is fixed, irrespectively of how we cluster them to construct the super-cells, the scaling law remains fixed. On the other hand, from Theorem 1 (MISO capacity scaling), the aggregate amount of information exchange is upper bounded by $MQN_t \lceil \log QN_t \rceil$. Once having a fixed number of base-stations in the network, i.e., MQ is fixed, the amount of information exchange is clearly minimized
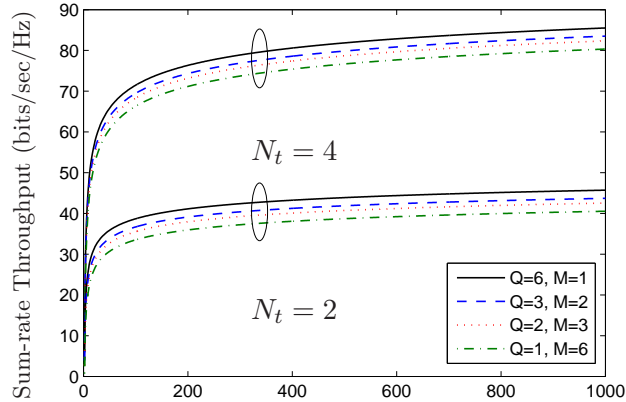
19

Figure 4: The impact of super-cell sizes Q on the achievable sum-rate throughput.

by setting $Q = 1$. This choice of Q corresponds to the setting that there is no cooperation among the base-stations and each base-station acts independently of others. This fact in conjunction with the freedom in clustering the network shows that even when the network is completely distributed (no base-station cooperation) by exchanging as low as $MN_t \log N_t$ information bits, still the optimal sum-rate throughput scaling is achievable.

There exists, however, a cost associated with such distributed downlink transmission. While the sum-rate throughput of the proposes scheme and that of dirty-paper coding scale identically, there exists a gap between them. The distributed processing penalizes the performance by enlarging this gap. By increasing the size of the super-cells Q we can reduce this gap, which shows that we can trade additional information exchange in favor of increasing the achievable sum-rate throughput, which is expected intuitively. The monte-carlo simulation results in Fig. 4 depicts the impact of clustering on the achievable sum-rate throughput. We consider a network containing $MQ = 6$ base-stations and look into the sum-rate of the four possible clustering cases of $(M, Q) = (6, 1), (3, 2), (2, 3)$ and $(1,6)$. While all four cases exhibit identical scalings, it is observed allowing more level of cooperation among the base-stations, i.e., large clusters sizes Q, improves the sum-rate throughput. For the simulations we have considered the two cases of $N_t = 1, 4$ and have assumed that the users have $N_r = 3$ receive antennas and $\rho = 10$ dB.

20

6.2 Fairness

In a non-homogenous network the reception quality at various users not only is influenced by how their channel direction vectors are aligned with the beam direction, but also is impacted by the path-loss and the shadowing they experience. As scheduling in our scheme involves opportunistic selection of the users who have satisfied a minimum level of reception quality, the system might be dominated by the user with the strongest reception qualities. In general, there exist an inherent tension between opportunistic user selection and maintaining some notion of fairness among the users. Therefore, it is imperative to investigate how such opportunistic user selection affects fairness.

It is, interestingly, observed that by deploying the proposed scheduling procedure, fairness, in the sense that all users are equiprobable to be served by the base-stations, is maintained. Such fairness is preserved as we assign different normalization factors to the users with different path-loss and shadowing ($\gamma_{m,n}^{k,r}$). This argument is formalized in the next theorem.

Theorem 4 (Fairness) *For downlink transmission in a MIMO network with M super-cells each with Q base-stations, K users per super-cell, N_t transmit antenna per base-station, and N_r receive antennas pre user, all users are equally likely to be served by the proposed scheduling procedure.*

Proof: The proof relies on the choices of the normalization factors given in (26). Let us define $\mathsf{SINR}_n^k(r, \ell, i)$ as received SINR of the ith receive antenna of user U_n^k when served along the beamforming vector $\boldsymbol{\omega}_n^r(\ell)$. By following the same line of argument as in (25) and taking into account (26), the probability that the ith receive antenna of user U_n^k is a candidate to be scheduled with the beamformer $\boldsymbol{\omega}_n^r(\ell)$ is

$$\theta_n \triangleq P\Big(\mathsf{SINR}_n^k(r, \ell, i) \geq \beta_n^k(r)\Big) = 1 - F_n^k(\beta_n^k(r); r) = \frac{1}{KN_r} \ , \tag{28}$$

which is identical for all users (in the nth super-cell) and all transmit and receive antennas. Therefore, the probability that exactly $j \leq \min(N_t, N_r)$ number of the receive antennas of user k being eligible to be served by a base-station is

$$\theta_n^j \triangleq \binom{N_t}{j} \theta_n^j (1 - \theta_n)^{N_t - j},$$

which is also identical for all users in the nth super-cell. Therefore, all users in each super-cell are equally likely to be scheduled for receiving any certain number of information streams. ∎

6.3 Training

As briefly mentioned at the end of Section 3 each user needs to calculate the SINRs that each of the beamformers can sustain for carrying the information intended for that user. Acquiring such SINRs is viable during the training interval. Such training being carried out on the downlink channel will cost downlink bandwidth. Nevertheless, it can be readily shown that such bandwidth loss does not penalize the throughput scaling rate. More specifically, due to the broadcast nature of the wireless channels, the base-stations will transmit a common training message for all users. The duration of such training will be *fixed* and independent of the number of users in the network. If we denote the fixed throughput loss due to such training by R_{train}, then the throughput after penalizing it by the training bandwidth loss will be $\mathbb{E}_{\boldsymbol{H}}[R] - R_{\mathsf{train}}$. R_{train} being fixed provides that for increasing K, $\mathbb{E}_{\boldsymbol{H}}[R] - R_{\mathsf{train}}$ will scale similar to $\mathbb{E}_{\boldsymbol{H}}[R]$. Therefore, while training incurs a loss on the throughput, it will not affect the throughput scaling rate provided in Theorem 2.

7 Conclusions

In this paper we have assessed the minimum amount of information exchange required in a multi-cell multiuser downlink transmission with base-station cooperation, in order to retain the optimal capacity scaling. By providing a constructive proof we have demonstrated that by an appropriate design of a feedback mechanism for each super-cell, we can avoid any cross-(super) cellular information exchange and yet achieve the optimal scaling law. More specifically, for a MIMO network consisting of M super-cells, each consisting of Q base-stations that carry out the downlink transmission collaboratively, when each base-station is equipped with N_t transmit antennas, exchanging $MQN_t\lceil \log QN_t \rceil$ information bits across the network suffices to guarantee the achievability of the optimal scaling of the sum-rate throughput. This result relies on a constructive proof, where we offer a scheduling framework that satisfies the information exchange constraints. Furthermore, the scheme also provides good performance for practical networks and sustain fairness among the users in the network.

A Proof of Lemma 1 (Conditional Expected Rates)

Scheme of the Proof:

$$\mathcal{H}_n^r(\ell) = \{k \mid \mathsf{sinr}_n^k(r,\ell) \geq 1\} \ . \tag{29}$$

This is due to the fact that when $\beta_n^k(r) \geq 1$, each user in the super-cell $\mathcal{Q}_n$ will have a normalized SINR greater than 1 for *at most* one of QN_t beamformers $\{\boldsymbol{\omega}_n^r(\ell)\}_{r,\ell}$, i.e., if a user has a normalized SINR ≥ 1 for a beamformer, its normalized SINR for the rest of beamformers is smaller than 1. This can be readily verified by noting that when $\beta_n^k(r) \geq 1$ the we have

$$\mathsf{sinr}_n^k(r,\ell) \geq 1 \quad \overset{(4)}{\Rightarrow} \quad \mathsf{SINR}_n^k(r,\ell) \geq \beta_n^k(r) \geq 1 \quad \overset{(2)}{\Rightarrow} \quad \mathsf{SINR}_n^k(r',\ell') < 1 \quad \forall (r',\ell') \neq (r,\ell)$$

$$\overset{(4)}{\Rightarrow} \quad \mathsf{sinr}_n^k(r',\ell') < 1 \quad \forall (r',\ell') \neq (r,\ell)$$

Given the above characterization of $\mathcal{H}_n^r(\ell)$ in (29), the rest of the proof has three major steps. In the first step by using the characterization of $\mathcal{H}_n^r(\ell)$ and some simple manipulation the conditional expected value of the rate $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ is reformulated. In the second step, we find a lower bound on this conditional expected rate. This lower bound itself is also a conditional expected rate with a different condition compared with $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$. Finally, in the third step we show how removing these new conditions from the conditional expected rates will provide a new lower bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ which is appropriate related to the order statistics of the set $\{\mathsf{SINR}_n^1(r,\ell), \ldots, \mathsf{SINR}_n^K(r,\ell)\}$.

Step 1) Reformulating $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$:
By recalling (9) we have

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] = \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \mathcal{H}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^k(r,\ell)\right) \mid \mathcal{H}_n^r(\ell)\right] . \tag{30}$$

On the other hand, from (29) we have

$$\forall k \in \mathcal{H}_n^r(\ell) : \mathsf{sinr}_n^k(r,\ell) \geq 1 \quad \text{and} \quad \forall k' \notin \mathcal{H}_n^r(\ell) : \mathsf{sinr}_n^{k'}(\ell) < 1 \ . \tag{31}$$

By combining (30) and (31) we find that

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$$

$$= \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \mathcal{H}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + \mathsf{SINR}_n^k(r,\ell) \right) \mid \forall k \in \mathcal{H}_n^r(\ell) : \mathsf{sinr}_n^k(r,\ell) \geq 1; \right.$$

$$\left. \text{and} \quad \forall k' \notin \mathcal{H}_n^r(\ell) : \mathsf{sinr}_n^{k'}(\ell) < 1 \right]$$

$$= \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \mathcal{H}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}} \left[\log \left(1 + \mathsf{SINR}_n^k(r,\ell) \right) \mid \forall k \in \mathcal{H}_n^r(\ell) : \mathsf{SINR}_n^k(r,\ell) \geq \beta_n^k(r); \right.$$

$$\left. \text{and} \quad \forall k' \notin \mathcal{H}_n^r(\ell) : \mathsf{SINR}_n^{k'}(\ell) < \beta_n^{k'}(r) \right], \quad (32)$$

where in the first equality we have replaced (31) into (30) and in the second equality we have invoked that $\mathsf{sinr}_n^k(r,\ell) = \frac{\mathsf{SINR}_n^k(r,\ell)}{\beta_n^k(r)}$.

Step 2) Lower Bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$:

In the next step we manipulate the conditions of the expected rate given in (32) and obtain a more tractable lower bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$. The core of such manipulation is given in the following lemma.

Lemma 6 (Conditional Expectation) *For a random variable X, increasing function $g(\cdot)$ and real values $b \geq a$ that belong to the support of X we have*

$$\mathbb{E}\Big[g(X) \mid X \geq b\Big] \geq \mathbb{E}\Big[g(X) \mid X \geq a\Big]. \quad (33)$$

Proof: Define $Y \triangleq g(X)$ and denote the CDFs of X and Y by $F_X(x)$ and $F_y(y)$, respectively.

$$\mathbb{E}\Big[g(X) \mid X \geq b\Big] = \int_y y \, d\big(F_{Y \mid X \geq b}(y)\big) = \int_0^1 F_{Y \mid X \geq b}^{-1}(u) \, du \ . \quad (34)$$

On the other hand, by recalling that $g(X)$ is monotonic in X

$$F_{Y \mid X \geq b}(y) = P(g(X) \leq y \mid X \geq b) = \frac{P(b \leq X \leq g^{-1}(y))}{P(b \leq X)} = \frac{F_X(g^{-1}(y)) - F_X(b)}{1 - F_X(b)}$$

$$\leq \frac{F_X(g^{-1}(y)) - F_X(a)}{1 - F_X(a)} = P(g(X) \leq y \mid X \geq a)$$

$$= F_{Y \mid X \geq a}(y) \ ,$$

which yields that

$$\forall u \in [0,1]: \quad F_{Y \mid X \geq b}^{-1}(u) \geq F_{Y \mid X \geq a}^{-1}(u) \ . \quad (35)$$

Equations (34) and (35) provide that

$$\mathbb{E}\Big[g(X) \mid X \geq b\Big] \geq \int_0^1 F_{Y \mid X \geq a}^{-1}(u)\, du = \int_y y\, d\big(F_{Y \mid X \geq a}(y)\big) = \mathbb{E}\Big[g(X) \mid X \geq a\Big] \; .$$

∎

Next we construct a hypothetical event under which the users in $\mathcal{H}_n^r(\ell)$ satisfy some other SINR constraints. For this purpose, we introduce the set of users $\tilde{\mathcal{H}}_n^r(\ell)$ corresponding to $\mathcal{H}_n^r(\ell)$ such that

$$\tilde{\mathcal{H}}_n^r(\ell) \triangleq \{k \mid \mathsf{SINR}_n^k(r,\ell) \geq \min_i \beta_n^i(r)\} \; . \tag{36}$$

Now consider a hypothetical channel condition for which the event $\tilde{\mathcal{H}}_n^r(\ell)$ includes the same set of users as the the event $\mathcal{H}_n^r(\ell)$ does. Under this hypothetical conditions, i.e., $\tilde{\mathcal{H}}_n^r(\ell) = \mathcal{H}_n^r(\ell)$, by recalling the characterization of $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ given in (32) we set $b = \beta_n^k(r)$ and $a = \min_i \beta_n^i(r)$ and replace the conditions $\mathsf{SINR}_n^k(r,\ell) \geq \beta_n^k(r)$ by $\mathsf{SINR}_n^k(r,\ell) \geq \min_i \beta_n^i(r)$. By noting that $b \geq a$ from Lemma 6 we obtain

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$$
$$\overset{(33)}{\geq} \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \tilde{\mathcal{H}}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}}\bigg[\log\Big(1 + \mathsf{SINR}_n^k(r,\ell)\Big) \mid \forall k \in \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^k(r,\ell) \geq \min_i \beta_n^i(r);$$
$$\text{and} \quad \forall k' \notin \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^{k'}(\ell) < \beta_n^{k'}(r)\bigg] \; . \tag{37}$$

Next, by taking into account the statistical independence of $\mathsf{SINR}_n^k(r;\ell)$ and $\mathsf{SINR}_n^{k'}(r;\ell)$ for $k \in \tilde{\mathcal{H}}_n^r(\ell)$ and $k' \notin \tilde{\mathcal{H}}_n^r(\ell)$, which is due to the statistical independence of channel matrices $\boldsymbol{H}_{n,m}^{k,r}$ and $\boldsymbol{H}_{n,m}^{k',r}$, changing the conditions enforced on $\mathsf{SINR}_n^{k'}(r;\ell)$ for $k' \notin \tilde{\mathcal{H}}_n^r(\ell)$ does not affect the statistics of $\mathsf{SINR}_n^k(\ell)$ for $k \in \tilde{\mathcal{H}}_n^r(\ell)$. Therefore, slightly changing the conditions $\mathsf{SINR}_n^{k'}(\ell) < \beta_n^{k'}(r)$ to $\mathsf{SINR}_n^{k'}(\ell) < \min_i \beta_n^i(r)$ for all $k' \notin \tilde{\mathcal{H}}_n^r(\ell)$ does not change the distribution of $\mathsf{SINR}_n^k(r,\ell)$ and (37) can be equivalently cast as

$$\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$$
$$\geq \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{k \in \tilde{\mathcal{H}}_n^r(\ell)} \mathbb{E}_{\boldsymbol{H}}\bigg[\log\Big(1 + \mathsf{SINR}_n^k(r,\ell)\Big) \mid \forall k \in \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^k(r,\ell) \geq \min_i \beta_n^i(r);$$
$$\text{and} \quad \forall k' \notin \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^{k'}(\ell) < \min_i \beta_n^i(r)\bigg] \; . \tag{38}$$

The new condition

$$\forall k \in \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^k(r,\ell) \geq \min_i \beta_n^i(r) \quad \text{and} \quad \forall k' \notin \tilde{\mathcal{H}}_n^r(\ell) : \mathsf{SINR}_n^{k'}(\ell) < \min_i \beta_n^i(r) \; ,$$

implies that under this condition, the set $\tilde{\mathcal{H}}_n^r(\ell)$ contains the $|\tilde{\mathcal{H}}_n^r(\ell)| = |\mathcal{H}_n^r(\ell)|$ largest values of the set

$$\{\mathsf{SINR}_n^1(r,\ell),\ldots,\mathsf{SINR}_n^K(r,\ell)\}\ ,$$

and the lower bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ given in (38) is the arithmetic average of these $|\mathcal{H}_n^r(\ell)|$ largest values conditioned on the constraint that they are larger than or equal to $\min_i \beta_n^i(r)$ and the remaining $(K - |\mathcal{H}_n^r(\ell)|)$ SINR values are smaller than $\min_i \beta_n^i(r)$. This interpretation can be formalized as follows.

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] \\
&\geq \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{j=1}^{|\mathcal{H}_n^r(\ell)|} \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r,\ell)\right) \;\middle|\; \mathsf{SINR}_n^{(A)}(r,\ell) \geq \min_i \beta_n^i(r) > \mathsf{SINR}_n^{(A+1)}(r,\ell) \right],
\end{aligned}
$$

$$(39)$$

where we have defined $A \overset{\triangle}{=} |\mathcal{H}_n^r(\ell)|$.

Step 3) Removing the Condition in (39)**:**

In the final step, we further simplify the lower bound on $\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)]$ provided in (39). This step will be the immediate application of the following lemma.

Lemma 7 (Conditional Expected Value of the Order Statistics) *Consider the set of independent random variables $\{X_1,\ldots,X_K\}$ and its corresponding set of order statistics $\{X^{(1)},\ldots,X^{(K)}\}$. For any increasing function $g(\cdot)$ and the real value a belonging to the supports of X_k we have*

$$\forall\, j \leq l: \quad \mathbb{E}\left[g\big(X^{(j)}\big) \mid X^{(l+1)} \leq a \leq X^{(l)}\right] \geq \mathbb{E}\left[g\big(X^{(j)}\big)\right]. \tag{40}$$

Proof: In the first step, by following the same line of argument as in the proof of Lemma 6 (conditional expectation) it can be readily shown that

$$\forall\, j \leq l: \quad \mathbb{E}\left[g\big(X^{(j)}\big) \mid X^{(l+1)} \leq a \leq X^{(l)}\right] \geq \mathbb{E}\left[g\big(X^{(j)}\big) \mid X^{(j+1)} \leq a \leq X^{(j)}\right]. \tag{41}$$

In the second step we also prove that

$$\forall\, j: \quad \mathbb{E}\left[g\big(X^{(j)}\big) \mid X^{(j+1)} \leq a \leq X^{(j)}\right] \geq \mathbb{E}\left[g\big(X^{(j)}\big)\right]. \tag{42}$$

Equations (41) and (42) together establish the desired result. For proving (42) we start by showing that for any real $b \geq a$

$$\forall j: \quad P\left(X^{(j)} \leq b \mid X^{(j+1)} \leq a \leq X^{(j)}\right) \leq P\left(X^{(j)} \leq b\right). \tag{43}$$

For this purpose let us define $\pi_i \triangleq (X_{i_1}, \ldots, X_{i_K})$ as the ith possible ordered permutations of the set $\{X_1, \ldots, X_K\}$ for $i = 1, \ldots, K!$. By averaging over all such permutations we have $\forall j$

$$P\left(X^{(j)} \leq b \mid X^{(j+1)} \leq a \leq X^{(j)}\right)$$

$$= \sum_i P\left(X^{(j)} \leq b \mid X^{(j+1)} \leq a \leq X^{(j)}, \ (X^{(1)}, \ldots, X^{(K)}) = \pi_i\right) P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= \sum_i P\left(X_{i_1}, \ldots, X_{i_{j-1}} \leq +\infty \text{ and } X_{i_j}, \ldots, X_{i_K} \leq b \mid X_{i_{j+1}}, \ldots, X_{i_K} \leq a \leq X_{i_1}, \ldots, X_{i_j}\right)$$

$$\times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= \sum_i \frac{P\left(\{a \leq X_{i_1}, \ldots, X_{i_{j-1}} \leq +\infty\} \text{ and } \{a \leq X_{i_j} \leq b\} \text{ and } \{X_{i_{j+1}}, \ldots, X_{i_K} \leq b\}\right)}{P\left(\{a \leq X_{i_1}, \ldots, X_{i_{j-1}} \leq +\infty\} \text{ and } \{a \leq X_{i_j}\} \text{ and } \{X_{i_{j+1}}, \ldots, X_{i_K} \leq b\}\right)}$$

$$\times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= \sum_i \frac{P\left(a \leq X_{i_j} \leq b\right)}{P\left(a \leq X_{i_j}\right)} \times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= \sum_i \frac{P\left(X_{i_j} \leq b\right) - P\left(X_{i_j} \leq a\right)}{P\left(a \leq X_{i_j}\right)} \times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$\leq \sum_i \frac{P\left(X_{i_j} \leq b\right) - P\left(X_{i_j} \leq a\right) P\left(X_{i_j} \leq b\right)}{P\left(a \leq X_{i_j}\right)} \times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= \sum_i P\left(X_{i_j} \leq b\right) \times P\left((X^{(1)}, \ldots, X^{(K)}) = \pi_i\right)$$

$$= P\left(X^{(j)} \leq b\right) ,$$

which establishes the claim in (43). By defining $Y^{(j)} \triangleq g(X^{(j)})$ and denoting the CDFs of $X^{(j)}$ and $Y^{(j)}$ by $F_{X^{(j)}}(x)$ and $F_{Y^{(j)}}(y)$, respectively, from (43) we immediately have

$$F_{Y^{(j)} \mid X^{(j+1)} \leq a \leq X^{(j)}}(y) \leq F_{Y^{(j)}}(y) \quad \Rightarrow \quad \forall u \in [0, 1]: \quad F^{-1}_{Y^{(j)} \mid X^{(j+1)} \leq a \leq X^{(j)}}(u) \geq F^{-1}_{Y^{(j)}}(u) . \quad (44)$$

Therefore, we get

$$\mathbb{E}\left[g(X^{(j)}) \mid X^{(j+1)} \leq a \leq X^{(j)}\right] = \int_y y \, d\left(F_{Y^{(j)} \mid X^{(j+1)} \leq a \leq X^{(j)}}(y)\right)$$

$$= \int_{u=0}^1 F^{-1}_{Y^{(j)} \mid X^{(j+1)} \leq a \leq X^{(j)}}(u) \, du$$

$$\overset{(44)}{\geq} \int_{u=0}^1 F^{-1}_{Y^{(j)}}(u) \, du = \int_y y \, d\left(F_{Y^{(j)}}(y)\right) = \mathbb{E}\left[g(X^{(j)})\right] ,$$

that justifies (42), which in turn along with (41) concludes the desired result. ∎

Now, by using the results of the lemmas above and by setting $X^{(j)} = \mathsf{SINR}_n^{(j)}(r, \ell)$ for $j \in \{1, \ldots, K\}$, $g(x) = \log(1 + x)$, and $l = A = |\mathcal{H}_n^r(\ell)|$, through invoking (40) (main result of Lemma 7) we obtain

$$
\forall j \leq A: \quad \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right) \mid \mathsf{SINR}_n^{(A)}(r, \ell) \geq \min_i \beta_n^i(r) > \mathsf{SINR}_n^{(A+1)}(r, \ell) \right]
$$
$$
\geq \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right) \right] . \tag{45}
$$

Combining (39) and (45) provides that

$$
\mathbb{E}_{\boldsymbol{H}}[R_n^r(\ell) \mid \mathcal{H}_n^r(\ell)] \geq \frac{1}{|\mathcal{H}_n^r(\ell)|} \sum_{j=1}^{|\mathcal{H}_n^r(\ell)|} \mathbb{E}_{\boldsymbol{H}}\left[\log\left(1 + \mathsf{SINR}_n^{(j)}(r, \ell)\right) \right] , \tag{46}
$$

which is the desired result.

B   Proof of Lemma 2 (SINR Bounds Distributions)

Based on these we define the following lower bound on $\mathsf{SINR}_n^k(r, \ell)$

$$
S_n^k(r, \ell) \triangleq \frac{|\boldsymbol{H}_{n,n}^k \boldsymbol{\omega}_n^\ell|^2}{\frac{\zeta_n^2}{\eta_n^1}\left(\sum_m \sum_l |\boldsymbol{H}_{n,m}^k \boldsymbol{\omega}_m^l|^2 - |\boldsymbol{H}_{n,n}^k \boldsymbol{\omega}_n^\ell|^2\right) + \frac{1}{\rho\eta_n^1}} , \tag{47}
$$

and the following upper bound on $\mathsf{SINR}_n^k(r, \ell)$

$$
T_n^k(r, \ell) \triangleq \frac{|\boldsymbol{H}_{n,n}^k \boldsymbol{\omega}_n^\ell|^2}{\frac{\zeta_n^1}{\eta_n^2}\left(\sum_m \sum_l |\boldsymbol{H}_{n,m}^k \boldsymbol{\omega}_m^l|^2 - |\boldsymbol{H}_{n,n}^k \boldsymbol{\omega}_n^\ell|^2\right) + \frac{1}{\rho\eta_n^2}} . \tag{48}
$$

Let $Y \triangleq |\boldsymbol{H}_{n,n}^{k,r} \boldsymbol{\omega}_n^r(\ell)|^2$. Since $\boldsymbol{H}_{n,n}^{k,r} \boldsymbol{\omega}_n^r(\ell) \sim \mathcal{CN}(0,1)$, Y has exponential distribution with unit variance. Also define $Z \triangleq \sum_m \sum_{r'} \sum_{\ell'} |\boldsymbol{H}_{n,n}^{k,r} \boldsymbol{\omega}_n^r(\ell)|^2 - |\boldsymbol{H}_{n,n}^k \boldsymbol{\omega}_n^\ell|^2$ which is the summation of $MQN_t - 1$ independent exponentially distributed random variables each with unit variance. Therefore, Z has a Gamma$(MQN_t - 1, 1)$ distribution. By denoting the probability density functions (PDF) of Z and Y by

$$
f_Y(y) = e^{-z} \quad \text{and} \quad f_Z(z) = \frac{z^{MQN_t - 2} e^{-z}}{(MQN_t - 2)!} ,
$$

the PDF of $T_n^k(r, \ell) = \frac{Y}{1/\rho\eta_n^2 + \frac{\zeta_n^1}{\eta_n^2} Z}$, denoted by $f_T(t)$ is

$$
f_T(t) = \int_0^\infty f_{T \mid Z}(t \mid z) f_Z(z) dz = \int_0^\infty \left(\frac{1}{\rho\eta_n^2} + \frac{\zeta_n^1}{\eta_n^2} z\right) e^{\left(-\frac{t}{\rho\eta_n^2} + \frac{\zeta_n^1}{\eta_n^2} tz\right)} \cdot \frac{z^{MQN_t - 2} e^{-z}}{(MQN_t - 2)!} dz
$$
$$
= \frac{e^{-t/\rho\eta_n^2}}{\left(\frac{\zeta_n^1}{\eta_n^2} t + 1\right)^{MQN_t}} \left[\frac{\frac{\zeta_n^1}{\eta_n^2} t + 1}{\rho\eta_n^2} + \frac{\zeta_n^1}{\eta_n^2}(MQN_t - 1) \right] ,
$$

where the last step holds as $\int_0^\infty e^{-u} u^M = M!$. Some simple manipulations the CDF of $T_n^k(r, \ell)$ is obtained as

$$F_2(x; n) = \int_0^x f_T(t) \, dt = 1 - \frac{e^{-x/\rho\eta_n^2}}{\left(\frac{\zeta_n^1}{\eta_n^2} x + 1\right)^{MQN_t - 1}} .$$

$F_1(x; n)$ can be found by following the same line of argument.

C   Proof of Lemma 3 (Ordered SINR Bounds)

By induction we show that for given n, r, ℓ, for all $j \in \{1, \dots, K\}$, we have $S_n^{(j)}(r, \ell) \leq \mathsf{SINR}_n^{(j)}(r, \ell)$.

1) For $j = 1$ we have

$$\mathsf{SINR}_n^{(1)}(r, \ell) = \max_k \mathsf{SINR}_n^k(r, \ell) \geq \max_k S_n^k(r, \ell) = S_n^{(1)}(r, \ell).$$

2) Assumption: For some $j = l$ we have $S_n^{(l)}(r, \ell) \leq \mathsf{SINR}_n^{(l)}(r, \ell)$.

3) Claim: For $j = l + 1$ we show that $S_n^{(l+1)}(r, \ell) \leq \mathsf{SINR}_n^{(l+1)}(r, \ell)$.

Based on the definition of $S_n^k(r, \ell)$, each of the $(K - l)$ terms $\{\mathsf{SINR}_n^{(l+1)}(r, \ell), \dots, \mathsf{SINR}_n^{(K)}(r, \ell)\}$ is greater than one corresponding element in the set $\{S_n^1(r, \ell), \dots, S_n^K(r, \ell)\}$. Therefore, there cannot be more than l elements in the set $\{S_n^1(r, \ell), \dots, S_n^K(r, \ell)\}$ that are all greater than all the members of $\{\mathsf{SINR}_n^{(l+1)}(r, \ell), \dots, \mathsf{SINR}_n^{(K)}(r, \ell)\}$. Now, by contradiction assume that $S_n^{(l+1)}(r, \ell) > \mathsf{SINR}_n^{(l+1)}(r, \ell)$. This assumption provides that

$$S_n^{(1)}(r, \ell) \geq S_n^{(2)}(r, \ell) \geq \cdots \geq S_n^{(l+1)}(r, \ell) > \mathsf{SINR}_n^{(l+1)}(r, \ell) \geq \mathsf{SINR}_n^{(l+2)}(r, \ell) \geq \mathsf{SINR}_n^{(K)}(r, \ell) ,$$

which indicates that we have found $l + 1$ members of the set $\{S_n^1(r, \ell), \dots, S_n^K(r, \ell)\}$ that are greater than all the $(K - l)$ members of the set $\{\mathsf{SINR}_n^{(l+1)}(r, \ell), \dots, \mathsf{SINR}_n^{(K)}(r, \ell)\}$ which contradicts with what we found earlier. Hence, we should have $S_n^{(l+1)}(r, \ell) \leq \mathsf{SINR}_n^{(l+1)}(r, \ell)$. The proof for the inequality $T_n^{(j)}(r, \ell) \geq \mathsf{SINR}_n^{(j)}(r, \ell)$ follows the same line of argument.

D   Proof of Lemma 4 (Bounds on $\bar{R}_n$)

The proof heavily depends on the result of the following lemma.

Lemma 8 *For a real variable $x \in [0, 1]$ and integer variables K and j, $0 \leq j \leq K - 1$, the function*

$$f(x, j) \triangleq \sum_{i=0}^{j} \binom{K}{i} x^{K-i} (1 - x)^i$$

is increasing in x.

Proof: See Appendix F. ∎

Note that from the properties of the order statistics [19] we know that

$$F_1^{(j)}(x;n) \triangleq \sum_{i=0}^{j-1} \binom{K}{i}\Big(F_1(x;n)\Big)^{K-i}\Big(1-F_1(x;n)\Big)^i \,, \tag{49}$$

$$F_2^{(j)}(x;n) \triangleq \sum_{i=0}^{j-1} \binom{K}{i}\Big(F_2(x;n)\Big)^{K-i}\Big(1-F_2(x;n)\Big)^i \tag{50}$$

$$G^{(j)}(x) \triangleq \sum_{i=0}^{j-1} \binom{K}{i}\Big(G(x)\Big)^{K-i}\Big(1-G(x)\Big)^i \,. \tag{51}$$

Also by referring to the definitions of $G(x)$ and $F_1(x;n)$ we find that

$$G\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)+\frac{x}{\rho\eta_n^1}\Big) \;=\; 1-\exp\Big[-\frac{x}{\rho\eta_n^1}-\underbrace{(MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)}_{\geq \ln\Big(1+\frac{\zeta_n^2}{\eta_n^1}x\Big)}\Big]$$

$$\geq\; 1-\frac{e^{-x/\rho\eta_n^1}}{\Big(1+\frac{\zeta_n^2}{\eta_n^1}x\Big)^{MQN_t-1}} = F_1(x;n) \,.$$

By applying Lemma 8, for $j=1,\ldots,K$, we have

$$G^{(j)}\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)+\frac{x}{\rho\eta_n^1}\Big) \;=\; f\Big(G\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)+\frac{x}{\rho\eta_n^1}\Big), j-1\Big)$$

$$\geq\; f\Big(F_1(x;n), j-1\Big)$$

$$=\; F_1^{(j)}(x;n) \,,$$

and consequently,

$$G^K\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)+\frac{x}{\rho\eta_n^1};n,r,\ell\Big) \;=\; \sum_{j=1}^{K} q_n^j(r,\ell)G^{(j)}\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+x)\Big)+\frac{x}{\rho\eta_n^1}\Big)$$

$$\geq\; \sum_{j=1}^{K} q_n^j(r,\ell)F_1^{(j)}(x;n) = F_1^K(x;n,r,\ell). \tag{52}$$

Define $u \triangleq F_1^K(x;n,r,\ell)$. Since $F_1^{(j)}(x;n)$, $j=1,\ldots,K$, are increasing functions in x, the function $F_1^K(x;n,r,\ell)$ is also increasing in x and thereof convertible. Therefore, (52) can be rewritten as $x=(F_1^K)^{-1}(u;n,r,\ell)$. By noting $G^K(x)$ is also invertible we have

$$G^K\Big((MQN_t-1)\Big(\frac{\zeta_n^2}{\eta_n^1}(1+(F_1^K)^{-1}(u;n,r,\ell))\Big)+\frac{(F_1^K)^{-1}(u;n,r,\ell)}{\rho\eta_n^1};\ell\Big) \geq u \,,$$

or equivalently

$$\Big(\rho\zeta_n^2(MQN_t-1)+1\Big)\Big(1+(F_1^K)^{-1}(u;n,\ell)\Big) \geq 1+\rho\eta_n^1(G^K)^{-1}(u;n,r,\ell) \,. \tag{53}$$

Using (53) we find

$$\sum_{r,\ell} \int_0^\infty \log(1+x)\, dF_1^K(x;n,r,\ell) + QN_t \log\left(\rho\zeta_n^2(MQN_t-1)+1\right)$$

$$= \sum_{r,\ell} \int_0^1 \log\left(1 + (F_1^K)^{-1}(u;n,r,\ell)\right) du + QN_t \log\left(\rho\zeta_n^2(MQN_t-1)+1\right) \int_0^1 du$$

$$= \sum_{r,\ell} \int_0^1 \log\left[\left(\rho\zeta_n^2(MQN_t-1)+1\right)\left(1 + (F_1^K)^{-1}(u;n,r,\ell)\right)\right] du$$

$$\geq \sum_{r,\ell} \int_0^1 \log\left(1 + \rho\eta_n^1(G^K)^{-1}(u;n,r,\ell)\right) du$$

$$= \sum_{r,\ell} \int_0^\infty \log(1+\rho\eta_n^1 x)\, dG^K(x;n,r,\ell) . \tag{54}$$

By substituting (54) into (21) we get the following lower bound on $\bar{R}_n$

$$\sum_{r,\ell} \int_0^\infty \log(1+\rho\gamma_1 x)\, dG^K(x;n,r,\ell) - QN_t \log\left(\rho\zeta_n^2(MQN_t-1)+1\right) \leq \bar{R}_n. \tag{55}$$

Also it can be easily verified that $\forall x \in \mathbb{R}$, $F_2(x;n) \geq G\left(\frac{x}{\rho\eta_n^2}\right)$. Hence,

$$F_2^K(x;n,r,\ell) = \sum_{j=1}^K q_j^n(r,\ell)\, F_2^{(j)}(x;n)$$

$$= \sum_{j=1}^K q_j^n(r,\ell)\, f\left(F_2(x;n), j-1\right)$$

$$\geq \sum_{j=1}^K q_j^n(r,\ell)\, f\left(G\left(\frac{x}{\rho\eta_n^2}\right), j-1\right)$$

$$= \sum_{j=1}^K q_j^n(r,\ell)\, G^{(j)}\left(\frac{x}{\rho\eta_n^2}\right) = G^K\left(\frac{x}{\rho\eta_n^2};n,r,\ell\right),$$

By defining $u \triangleq \frac{F_2^K(x;n,\ell)}{\rho\eta_n^2}$ and following the same lines as above we ge

$$\sum_{r,\ell} \int_0^\infty \log(1+x)\, dF_2^K(x;n,r,\ell) \leq \sum_{r,\ell} \int_0^\infty \log(1+\rho\gamma_2 x)\, dG^K(x;n,r,\ell) . \tag{56}$$

Inequalities in (54) and (56) together give rise to the following upper bound on $\bar{R}_n$

$$\bar{R}_n \leq \sum_{r,\ell} \int_0^\infty \log(1+\rho\gamma_2 x)\, dG^K(x;n,r,\ell) . \tag{57}$$

Combining the lower and upper bounds given in (55) and (57) establishes the desired result.

E   Proof of Lemma 5 (Scaling)

We start by citing the following theorem.

Theorem 5 [13, Theorem 4] *Let $\{X_n\}_{n=1}^{K}$ be a family of positive random variables with finite mean μ_K and variance σ_K^2, also $\mu_K \to \infty$ and $\frac{\sigma_K}{\mu_K} \to 0$ as $K \to \infty$. Then, for all $\alpha \in \mathbb{R}^+$ we have*

$$\mathbb{E}\left[\log(1 + \alpha X_K)\right] \doteq \log\left(1 + \alpha\mathbb{E}[X_K]\right).$$

Now, for any given number of users K per super-cell, we define a random variable X_K that is distributed as $X_K \sim W^K(x)$. Also for $j = 1, \ldots, K$ we define

$$\mu_{(j)} \triangleq \int_0^\infty x \, dW^{(j)}(x) \,,$$

$$\text{and} \quad \sigma_{(j)}^2 \triangleq \int_0^\infty \left(x - \mu_{(i)}\right)^2 dW^{(j)}(x) \,,$$

$$\text{and} \quad \mu_K \triangleq \mathbb{E}[X_K] = \int_0^\infty x \, dW^K(x)$$

$$= \sum_{j=1}^{K} Q_j \int_0^\infty x \, dW^{(j)}(x) = \sum_{j=1}^{K} Q_j \mu_{(j)} \,.$$

As discussed in [19, 12, Sec. 4.6], for ordered exponentially distributed random variables X_K we have

$$\sigma_K^2 < 2 + 2\mu_{(1)}\left(\mu_{(1)} - \mu_K\right) \,, \tag{58}$$

$$\text{and} \quad \log K + \zeta + \frac{1}{2(K+1)} \leq \mu_{(1)} \leq \log K + \zeta + \frac{1}{2K} \,, \tag{59}$$

where by noting that $\zeta \approx 0.577$ is the Euler-Mascheroni constant we have $\mu_{(1)} \doteq \log K$. Also

$$\mu_{(1)} - \log\left(\sum_{j=1}^{K} jQ_j\right) - \zeta - 0.5 \leq \mu_K \leq \mu_{(1)}.$$

By taking into account the constraint in (23), as $K \to \infty$

$$1 - \frac{\log\left(\sum_{j=1}^{K} jQ_j\right) - \zeta - 0.5}{\log K} \leq \mu_K \leq 1. \tag{60}$$

Equations (59) and (60) together show that

$$\mu_{(1)} \doteq \mu_K \doteq \log K, \tag{61}$$

which also implies that $\mu_K \to \infty$. Taking into account (58) and (61) we also conclude that $\lim_{K\to\infty} \frac{\sigma_K}{\mu_K} = 0$ and therefore the conditions of Theorem 5 are satisfied. Hence, from Theorem 5

$$
\begin{aligned}
\int_0^\infty \log(1 + ax)\, dW^K(x) &= \mathbb{E}\Big[\log(1 + aX_K)\Big] \doteq \log\Big(1 + a\mathbb{E}[X_K]\Big) \\
&= \log\Big(1 + a\mu_K\Big) \doteq \log\log K + \log(a) .
\end{aligned}
$$

F   Proof of Lemma 8

By the expansion of $\Big(x + (1 - x)\Big)^K$ we have

$$
\begin{aligned}
f(x, j) &= 1 - \sum_{i=j+1}^{K} \binom{K}{i} x^{K-i}(1 - x)^i = 1 - \sum_{i=j+1}^{K} \binom{K}{K-i} x^{K-i}(1 - x)^i \\
&= 1 - \sum_{i=0}^{K-(j+1)} \binom{K}{i} (1 - x)^{K-i} x^i = 1 - f(1 - x, K - j - 1),
\end{aligned}
$$

where it can be concluded that $f'(x, i) = f'(1 - x, K - j - 1)$. So it is sufficient to show that $f'(x, i) \geq 0$ for $x \leq \frac{1}{2}$ and for all $j = 1, \ldots, K - 1$. For this purpose we consider two cases of $j \leq \lfloor \frac{K}{2} \rfloor$ and $j > \lfloor \frac{K}{2} \rfloor$.

Case 1: $j \leq \lfloor \frac{K}{2} \rfloor$

$$
\begin{aligned}
f'(x, j) &= \sum_{i=0}^{j} \binom{K}{i} (K - i) x^{K-i-1}(1 - x)^i - \binom{K}{i} i x^{K-i}(1 - x)^{i-1} \\
&= \sum_{i=0}^{j} \binom{K}{i} (K - i) x^{K-i-1}(1 - x)^{i-1} \Big[K(1 - x) - i \Big],
\end{aligned}
\tag{62}
$$

where it can be shown that for $x \leq \frac{1}{2}$

$$
K(1 - x) - i \geq K(1 - x) - j \geq K(1 - x) - \frac{K}{2} = \frac{K}{2}(1 - 2x) \geq 0.
$$

Case 2: $j > \lfloor \frac{K}{2} \rfloor$

Define $a_i = 1 - \frac{1}{2}\delta(\lfloor \frac{K}{2} \rfloor - i)$, where $\delta(\cdot)$ is the Dirac delta function. Therefore, we get

$$
f(x, j) = f(x, K - j - 1) + \sum_{i=K-j}^{\lfloor \frac{K}{2} \rfloor} a_i \binom{K}{i} \Big[x^{K-i}(1 - x)^i + x^i(1 - x)^{K-i} \Big].
$$

For $x \leq \frac{1}{2}$ we get

$$
\begin{aligned}
f'(x,j) &= f'(x, K-j-1) \\
&+ \sum_{i=K-i}^{\lfloor \frac{K}{2} \rfloor} a_i \binom{K}{i} x^{K-i-1}(1-x)^{i-1}\Big[K-i-Kx\Big] + \underbrace{x^{i-1}(1-x)^{K-i-1}}_{\geq x^{K-i-1}(1-x)^{i-1}}\Big[i-Kx\Big] \\
&\geq f'(x,\underbrace{K-j-1}_{\leq \lfloor \frac{K}{2} \rfloor}) + \sum_{i=K-i}^{\lfloor \frac{K}{2} \rfloor} a_i \binom{K}{i} x^{K-i-1}(1-x)^{i-1}\underbrace{\Big[K-2Kx\Big]}_{\geq 0} \geq 0 \quad (63)
\end{aligned}
$$

From (62) and (63) it is concluded that for $x \leq \frac{1}{2}$, $f(x,j)$ is an increasing function of x, which provides the proof.

References

[1] G. Caire and S. Shamai, "On the achievable throughput of a multiantenna Gaussian broadcast channel," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1691– 1706, Jul. 2003.

[2] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 5, pp. 684–702, Jun. 2003.

[3] M. Sharif and B. Hassibi, "A comparison of time-sharing, DPC, and beamforming for MIMO broadcast channels with many users," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 11–15, Jan. 2007.

[4] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Slzer, "Shifting the MIMO paradigm," *IEEE Signal Process. Mag.*, vol. 24, no. 5, pp. 36–46, Sept. 2007.

[5] H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and H. Dai, "Asynchronous interference mitigation in cooperative base station systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 1, pp. 155–165, Jan. 2008.

[6] H. Huang, M. Trivellato, A. Hottinen, M. Shafi, P. Smith, and R. Valenzuela, "Increasing downlink cellular throughput with limited network MIMO coordination," *IEEE Trans. Wireless Commun.*, vol. 8, no. 6, pp. 2983–2989, June 2009.

[7] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath, "Networked MIMO with clustered linear precoding," *IEEE Trans. Wireless Commun.*, vol. 8, no. 4, pp. 1910–1921, April 2009.

[8] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 506–522, Feb. 2005.

[9] T. Yoo, N. Jindal, and A. Goldsmith, "Multi-antenna downlink channels with limited feedback and user selection," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 7, pp. 1478–1491, Sep. 2007.

[10] K. Huang, R. W. Heath, and J. G. Andrews, "Space division multiple access with a sum feedback rate constraint," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3879–3891, Jul. 2007.

[11] K. Huang, J. G. Andrews, and R. W. Heath, "Performance of orthogonal beamforming for SDMA with limited feedback," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 152–164, Feb. 2009.

[12] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, pp. 4363–4372, Nov. 2007.

[13] ——, "Opportunistic beamforming with limited feedback," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2765–2771, Aug. 2007.

[14] T. Tang and R. W. Heath, "Opportunistic feedback for downlink multiuser diversity," *IEEE Commun. Lett.*, vol. 9, no. 10, pp. 948–950, Oct. 2005.

[15] T. Tang, R. W. Heath, S. Cho, and S. Yun, "Opportunistic feedback in multiuser MIMO systems with linear receivers," *IEEE Trans. Commun.*, vol. 55, no. 5, pp. 1020–1032, May 2007.

[16] W.-Y. Shin, S.-W. Jeon, N. Devroye, M. H. Vu, S.-Y. Chung, Y. H. Lee, and V. Tarokh, "Improved capacity scaling in wireless networks with infrastructure," *IEEE Trans. Inf. Theory*, 2008, submitted, available at http://arxiv.org/pdf/0811.0726.

[17] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE International Conference on Computer Communications (INFOCOM)*, San Francisco, CA, April 2003, pp. 1543 – 1552.

[18] B. Hassibi and T. L. Marzetta, "Multiple-antennas and isotropically random unitary inputs: the received signal density in closed form," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1473–1484, Jun. 2002.

[19] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics.* New York: Wiley, 1992.